



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

Ancient DNA studies: new perspectives on old samples

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Ancient DNA studies: new perspectives on old samples / Ermanno Rizzi;Martina Lari;Elena Gigli;Gianluca De Bellis;David Caramelli. - In: GENETICS SELECTION EVOLUTION. - ISSN 1297-9686. - ELETTRONICO. - 44:(2012), pp. 21-40. [10.1186/1297-9686-44-21]

Availability:

This version is available at: 2158/674926 since:

Published version:

DOI: 10.1186/1297-9686-44-21

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

(Article begins on next page)

REVIEW

Open Access

Ancient DNA studies: new perspectives on old samples

Ermanno Rizzi¹, Martina Lari², Elena Gigli², Gianluca De Bellis¹ and David Caramelli^{2*}

Abstract

In spite of past controversies, the field of ancient DNA is now a reliable research area due to recent methodological improvements. A series of recent large-scale studies have revealed the true potential of ancient DNA samples to study the processes of evolution and to test models and assumptions commonly used to reconstruct patterns of evolution and to analyze population genetics and palaeoecological changes. Recent advances in DNA technologies, such as next-generation sequencing make it possible to recover DNA information from archaeological and paleontological remains allowing us to go back in time and study the genetic relationships between extinct organisms and their contemporary relatives. With the next-generation sequencing methodologies, DNA sequences can be retrieved even from samples (for example human remains) for which the technical pitfalls of classical methodologies required stringent criteria to guaranty the reliability of the results. In this paper, we review the methodologies applied to ancient DNA analysis and the perspectives that next-generation sequencing applications provide in this field.

Review

Twenty-eight years of ancient DNA

The field of ancient DNA studies began twenty-eight years ago with the extraction and sequencing of DNA material from the quagga, a South African equid (*Equus quagga quagga*) that went extinct in the 19th century [1] and from an Egyptian mummy [2]. These studies used bacterial cloning to amplify small DNA sequences (retrieved from skin fragments of these specimens) and showed that the origin of the extracted genetic material was mainly microbial or fungal. In general, endogenous DNA was composed of very low concentrations of short and damaged fragments of multi-copy loci, such as mitochondrial DNA (mtDNA).

A few years later, with the development of PCR (polymerase chain reaction) it became possible to routinely amplify and study surviving ancient DNA molecules even if only in a single copy, resulting in a rapid increase and diversification of ancient DNA research [3-5]. However, due to the enormous power of PCR to amplify even a few copies of DNA sequences, modern DNA contamination

has become a crucial problem. For this reason, many of the most extravagant reports on ancient DNA, including claims of DNA sequences surviving for millions of years in plants [6-8] and dinosaur bones [9], have been disputed and actually disregarded. Studies on ancient DNA need to deal with technical problems that are specific to this field. The first difficulty is the production of sufficient quantities of authentic DNA sequences to make a study conclusive. This difficulty is a consequence of post-mortem DNA degradation processes, which can cause miscoding lesions, potentially leading to sequence errors, or physical destruction of the DNA molecule, thus increasing the risk for preferential amplification of exogenous contaminant sequences (Table 1). To deal with this issue, researchers have outlined a series of guidelines to ensure the quality of ancient DNA data and the reliability of consequent conclusions [5]. Over the years, these guidelines have gradually evolved into a more detailed and extensive list of requirements, resulting in the nine "gold criteria" outlined by Cooper and Poinar in "Ancient DNA: do it right or not at all" [10], one of the most cited papers in the field (Additional file 1). In particular, the authors have suggested that, in the absence of full compliance with all nine criteria, the reliability and authenticity of results remain uncertain.

In the last few years, with the advent of new sequencing technologies, the field of ancient DNA is experiencing a

* Correspondence: david.caramelli@unifi.it

²Department of Evolutionary Biology, Laboratory of Anthropology, Molecular Anthropology/Paleogenetics Unit, University of Florence, Via del Proconsolo 12, Florence 50122, Italy

Full list of author information is available at the end of the article

Table 1 Ancient DNA damage

Damage	Type of process	Effects on DNA molecule	Possible solutions in aDNA classical sequencing methodologies
Oxidative damage	Formation of strand breaks (single-stranded nicks)	Cleavage of the phosphodiester backbone	PCR of overlapping fragments of short length
		Depurination resulting in a baseless site	Multiple independent PCR Cloning and sequencing of several clones
		Breakage of the sugar backbone through β -elimination	Uracil-N-glycosylase treatment
		Results in lesions blocking the polymerase enzyme, and promoting chimeric sequences through 'jumping' PCR	Blocking primers Single primer extension or Spex
Degradation by microorganisms' nucleases in the post mortem cell	Strand breaks	Short fragment length	PCR of overlapping fragments of short length
DNA crosslinks	Inter-strand crosslinks by alkylation	May prevent the amplification of endogenous template molecules	PTB (N-phenylacetyl thiazolium bromide)
	Intermolecular crosslinks by Maillard reaction	Increases the risk of contamination	
Hydrolysis damage	Results in miscoding lesions, for example, deamination of cytosine and adenine to uracil and hypoxanthine, respectively	Results in the incorporation of erroneous bases during amplification and change of coding	Multiple independent PCR Cloning and sequencing of several clones UNG treatment

UNG : Uracil-N-glycosylase; Spex: Single Primer Extension; PTB: N-phenylacetyl thiazolium bromide.

new era wherein what was once impossible is now possible, as for example drafting genomes of extinct organisms such as *Homo neanderthalensis* [11], or distinguishing endogenous from contaminant DNA in archaic *Homo sapiens* specimens [12]. In this report, we review the history of studies on ancient DNA from a methodological point of view, ranging from the most significant ones performed with the so-called "classical methodology" consisting in PCR amplification, cloning and traditional Sanger sequencing, to the more recent ones performed by next-generation sequencing technologies (NGS), which are promising to revolutionize the field of ancient DNA.

What was achieved with the classical methodology?

Starting from the first pioneering studies in the 1980s, classical methods to recover and analyze DNA from ancient specimens have been developed, and continuously improved, to overcome the two main technical

limitations characterizing this field, e.g. the poor preservation of endogenous DNA and the presence of contaminant exogenous DNA.

Basically, the traditional methodology consists in three fundamental steps:

1. PCR amplification of several short and overlapping target fragments (60–200 bp long) to recover larger regions.
2. Production and sequencing of several clones for each amplified fragment.
3. Alignment and comparison of sequences from different clones and different overlapping fragments to reconstruct the final consensus sequence of the entire region of interest.

Using this protocol, Krings et al. [13] have reported in 1997, the first reconstruction of a DNA sequence from an extinct hominin, *Homo sapiens neanderthalensis*. This

consisted in a small fragment of mitochondrial DNA (mtDNA) [13] that was subsequently enriched by other mtDNA sequences from the same individual [14]. A few years later, these results were corroborated by additional mtDNA sequences from other Neanderthal specimens distributed all over Europe [15-19].

Until recently, almost all the genetic studies performed on ancient specimens targeted mtDNA regions. Cells contain 100 to 10 000 copies of mtDNA that constitute a primary source of DNA in ancient specimens. In addition, the mitochondrial genome being maternally inherited and not subjected to recombination, mtDNA mutations are transmitted clonally across generations and can be used to trace maternal lineages. Because of these characteristics, mtDNA has been successfully used in several studies to reconstruct the phylogenetic relationships between extant and extinct species, such as Australian marsupial wolves [5,20], New Zealand moas [21,22], American ground sloths [23] endemic Hawaiian goose [24], cave bears [25], Balearic Islands cave goats [26], giant lemurs [27] and Caspian tigers [28].

The preservation of many individuals, all originating from a single locality, either as museum specimens previously collected by naturalists, or specimens retrieved by archaeologists at a single site, provides an opportunity to track changes in a population over time. The first example of this approach is a study on three populations of kangaroo rats in California, which were collected by zoologists in the first half of the last century. Comparison of modern populations from the same area has shown that the mitochondrial lineages are spatially stable, a situation typical of undisturbed habitats [29]. A more recent study conducted on mice sampled in Chicago (USA) has demonstrated that, in this population, mitochondrial lineages have been replaced over the last 150 years, probably due to human influence [30]. Similar studies have traced the population history of many other species including rabbits [31], pocket gophers [32], black-footed ferrets [33], sea otters [34], grizzlies [35], red squirrels [36], canids [37], penguins [38], arctic fox [39], Chatham Island taikos [40], woolly mammoths [41], bears [42], equids [1,43], Tasman booby [44] and quails [45]. Ramakrishnan and Hadly [46] have reviewed these studies and reanalyzed some datasets. They concluded that genetic samples collected from populations over large temporal and geographical scales, when analyzed using complex models and the serial coalescent approach, are critical to understand past population dynamics and provide important material to reconstruct evolutionary processes.

An interesting study on late Pleistocene brown bears has radically revolutionized the current knowledge about bear population dynamics in Alaska [47]. Whereas present-day brown bear mitochondrial lineages are

neatly distributed in different geographical areas, this study has shown that, about 35 000 years ago, the same mitochondrial lineages all coexisted in a single area. This result has crucial implications for conservation genetics. Mitochondrial lineages that are spatially separated today are often considered to have separated a long time ago and to be carried by "subspecies" adapted to different environments. As a consequence, it is often suggested that they should be kept separated and not be allowed to mix in captivity or during the restocking of wild species. The data on ancient DNA sequences from brown bear prove that contemporary samples may not reproduce long-term patterns. In the future, it can be predicted that direct testing of the phylogeographic patterns of additional species will clarify whether or not these results reflect recent effects of random genetic drift within small populations or, whether they represent long-term separation of populations.

Studies on ancient DNA have also been used to investigate the demographic history of human populations although such studies may suffer from contamination problems. One of the earliest publications in this field is a population genetics study on the Etruscans, a prehistoric Italian population [48]. Other studies have examined the origins of Andaman Islanders [49], the genetic variations in prehistoric Iberians and Sardinians [50,51], and the genealogical discontinuities among Etruscans, Medieval and contemporary Tuscans [52]. Particular attention has also been paid to the investigation of the genetic relationship between local hunter-gatherers and the first farmers during the Palaeolithic-Neolithic transition in Europe [53-55]. Although based on reduced sample sets, these studies revealed a genetic discontinuity between these populations.

Other studies on ancient DNA have aimed at tracing the genetic history of domestication. The most extensively studied species are cattle and pigs. The limited geographical range of sheep and goat wild ancestors indicates that the domestication of these species probably occurred in Anatolia and in the Fertile Crescent. On the contrary, the wild progenitors of cattle and pigs are distributed on a wider area indicating that their domestication scenario may be more complex with possible genetic contributions from local wild stocks.

The first report on ancient bovine DNA concerns the mtDNA sequences of the extinct wild oxen (*Bos primigenius*) from Britain [56,57]. Phylogenetic analysis of *Bos primigenius* ancient sequences has shown that these animals are clearly distinct from modern cattle. For this reason, they have been assigned to a different clade, named "P", in contrast to the "T" clade corresponding to modern domestic cattle. Even if based on only a few sequences from a single mtDNA region, the data point to a Near-Eastern origin of European cattle. More recent

genetic studies on aurochs and ancient domestic cattle from different locations have suggested that the process of cattle domestication may be more complex than previously thought [58-62]. When analyzing mtDNA from Pleistocene Italian aurochs, Beja-Pereira et al. [59] and Mona et al. [62], have found that the vast majority of the Italian sequences fall within the range of variation observed in modern cattle ("T" clade) and are genetically distinct from the "P" clade of the aurochs from the British Isles [62]. In 2007, Edwards et al. have reported that only "P" type sequences are present in ancient aurochs from North and Central Europe [61], while nearly all the sequences of the first domestic cattle of the Neolithic and Bronze Ages belong to the "T" clade [58,60]. These data represent good examples of how studies on ancient DNA can contribute to shed light on complex scenarios characterizing cattle domestication and diffusion processes. Indeed, populations domesticated in the Near East and introduced into Europe during the Neolithic diffusion might have intermixed in Southern European regions with local wild animals carrying mtDNA of the "T" type. In addition, a greater genetic variability has been observed in aurochs from South Europe than in those from North and Central Europe. Previously unknown non-"T" mtDNA sequences have been discovered in some modern local Italian breeds [63-65]. As a consequence, European breeds could represent a more diverse and important genetic resource than previously recognized, especially in Southern regions.

Two important papers by Larson et al. [66,67] based on ancient mtDNA sequences of both wild and domestic pigs from different times and locations, have provided insights into pig domestication processes and Neolithic expansion in Europe, as well as in East Asia and Oceania. By analyzing several ancient pig specimens collected across West Eurasia, the authors have shown that Near Eastern pigs have been introduced by humans into Europe and that they may have traveled along at least two distinct routes (the Danubian corridor and a southern maritime route that ran through the North shore of the Mediterranean) [67]. Then, European wild boars were domesticated and rapidly became the predominant lineage within the European domestic swine. In the light of these results, the process of pig domestication in Europe appears fundamentally different from that in the Near East. In Europe, it may not have been a truly independent event, but rather a direct consequence of the introduction of Near Eastern domestic pigs by early farmers. Other similar studies have investigated the domestication process of chicken and horse [68,69].

Improvements of the classical methodology

Since, even when preserved in ideal conditions, DNA degrades over time, the process of amplification, cloning

and sequencing is usually constrained by the short length of the DNA fragments recovered and by the small amount of template DNA in the sample. Although longer sequences are more informative, in practice, with the classical methodology, only overlapping shorter sequences are available that must then be meticulously assembled together. For this reason, until 2005 it was not possible to obtain DNA sequences longer than 1000 base pairs (bp), even from widely studied Pleistocene mammalian species such as mammoths, ground sloths and cave bears. With the breakthrough in ancient DNA sequencing due to the innovative multiplexing strategy of Krause et al. [70] to reconstruct long DNA sequences from several shorter fragments, the complete sequence of the 16 770 bp mitochondrial genome of the Pleistocene woolly mammoth (*Mammuthus primigenius*) was reported, starting from only about 200 mg of bone. Basically, the multiplexing strategy consists of a two-stage PCR: (1) first, multiple primer pairs are used to target subsequences within the complete DNA sequence and (2) the amplification product is divided in as many aliquots as the number of primer pairs used in the multiplex PCR and each aliquot serves as template for individual secondary PCR that separately amplify each target. For the woolly mammoth mtDNA, 46 such primer pairs were chosen to amplify overlapping DNA sequences covering the entire mtDNA genome. Then, primer pairs were divided into two sets, each one comprising primers amplifying alternate non-contiguous fragments. Multiplex PCR was then performed with only as much ancient DNA template as would be normally used for short target sequencing. Even if the multiplexing approach does not save time (many secondary PCR are still required), it certainly saves DNA material, since it can use hundreds of primer pairs in the same tube, thus overcoming typical limitations in sample amount.

Modifications have been made to this classical methodology in order to explore the molecular nature of the DNA damages occurring in such specimens. One common feature of ancient DNA samples is the presence of miscoding lesions that cause the incorporation of incorrect nucleotides during DNA amplification. In 2001, Hofreiter et al. showed for the first time that most of these damage-derived errors are caused by hydrolytic deamination of cytosine into uracil leading to apparent C → T or G → A substitutions in DNA templates sequenced after PCR amplification [71]. To reduce the number of such misincorporations, assays involving pre-treatment of ancient DNA samples with uracil N-glycosylase have been conducted to remove uracil residues from the DNA sequence and leave abasic sites that prevent replication by the Taq polymerase during PCR. However, the use of uracil N-glycosylase has not been widely adopted in studies with ancient DNA, because the number of template

molecules in such samples is often so small that the enzyme may destroy all the amplifiable templates. In 2007, Cooper et al. reported another approach, i.e. Single Primer Extension (SPEX), which can provide detailed information about post-mortem base modifications in ancient DNA [72]. Unlike PCR, SPEX is an amplification methodology that uses single biotinylated primers to specifically target only one strand of the ancient DNA template at a locus of interest without imposing any predefined target length. With SPEX, primer extension continues until it is stopped by the end of the template due to fragmentation or by a polymerase-blocking lesion. After the primer extension step, all non-biotinylated DNA templates are removed by washes in the presence of streptavidin beads and the remaining first-generation copies of the ancient DNA target are then ligated to a polyC-tail and amplified by nested PCR. Comparison of data obtained with SPEX amplification and standard procedures on the same samples showed that SPEX can produce sequence data of unprecedented accuracy from ancient DNA, without introducing additional exogenous sequence artifacts. Cooper et al. [72] also suggested that the C → U base modifications may originate from the damage-derived miscoding lesions present at significant levels in samples of ancient DNA. Despite its potential interest, after this initial report, SPEX has not been widely applied mainly because of its intrinsic laborious protocol compared to other more innovative procedures (see “Next Generation Sequencing methodology” section).

Another recent implementation in the classical methodology is the use of blocking oligonucleotides that preferentially bind modern contaminant DNA and prevent its amplification, in combination with standard primers specific for the target of interest. A blocking oligonucleotide consists in modifying its 3' end with a C3 spacer to prevent the TaqDNA polymerase from extending it without significantly changing its annealing properties to the target DNA. The C3 spacer is a standard primer modification available from most suppliers of custom oligonucleotides and consists in a short three-carbon spacer arm [73]. This novel PCR method has been tested by Gigli et al. on four Neanderthal samples with different contamination levels and taphonomic conditions [74]. Usually, Neanderthal skeletal remains are contaminated with modern human DNA derived from the handling and the washing of the specimens during excavation. Although Neanderthal and modern-human haplotypes differ, allowing the design of specific primer pairs for most of the mtDNA hypervariable region 1 (HVR1), the human contaminants can often outnumber the endogenous DNA, thus preventing a successful retrieval of Neanderthal sequences. However, including blocking oligonucleotides, specifically designed to preferentially bind modern-human specific sequences from

contaminant sources, has made it possible to significantly increase the yield of Neanderthal DNA sequences in all four samples (from 25.23 % up to 90.18 %) by inhibiting the hybridization of the standard primers.

Nuclear ancient DNA

Even if a sample of ancient DNA, typically contains 1000 times more copies of mtDNA than nuclear DNA, the number of mtDNA copies, sufficiently well preserved to be analyzed, is very low. In fact, nuclear DNA seems less prone to degradation and damage over time, so the chance of recovering longer intact strands is actually greater for nuclear DNA than for mtDNA. In addition, DNA damages are less frequent in nuclear DNA than in mtDNA, possibly because nuclear DNA is better protected by proteins.

From a quantitative point of view, DNA sequences that occur in many hundreds of copies per cell, such as DNA from mitochondria or chloroplasts in plants, are more easily retrieved from ancient specimens than are nuclear DNA sequences that occur only once per haploid genome. Therefore, phylogenies are usually reconstituted with information from only a few genetic loci, which limits phylogeny reconstruction of species that have diverged either recently or so rapidly that different parts of their genome have different phylogenies. However, this limitation can be overcome in some cases. For example, it has been possible to determine nuclear DNA sequences from several Pleistocene animals [75,76], from plants preserved in dry environments [77] and from museum specimens [78,79]. An interesting study on ancient cattle samples has investigated cattle domestication based on paternally transmitted Y chromosome-specific single nucleotide polymorphisms (SNP) [80,81]. In modern European cattle, one SNP in intron 19 of the Y-chromosomal gene *UTY* and one 2-bp insertion-deletion polymorphism in the *ZFY* intron 5, distinguish two haplogroups: Y1, primarily found in Northern Europe, and Y2, mainly distributed in Southern Europe [80]. Detection of the Y1 haplogroup in ancient cattle and aurochs from Northern and Central Europe has been initially interpreted as proof of a hybridization event having occurred in Europe between domestic cows and aurochs bulls [80]. However, subsequent studies on different sample sets have failed to find the same pattern, and shown a great fluctuation of allele frequencies over time [81,82]. These results suggest that the present-day frequencies of Y1 and Y2 haplogroups probably depend more on recent demographic events than on events having occurred in the first stages of cattle domestication. Even if the analysis of Y chromosome lineages is potentially useful to understand the evolutionary history of cattle, there is still no patrilineal marker capable of tagging a possible introgression of European aurochs into domestic cattle.

Having access to nuclear ancient DNA might enable scientists to investigate ancient animal and human phenotypes, such as skin color and behavioral traits. Of all the important questions on Neanderthals, that of whether they were able to speak and, if so, how well, is a key issue. To address this, Krause et al. [83] have analyzed the *FOXP2* gene, a gene intimately connected with the ability to speak in humans. They have applied the classical multiplexing approach, to characterize two SNP in the *FOXP2* gene and to check a number of modern human DNA contamination control positions on autosomes, sex chromosomes and mtDNA. They have found that Neanderthal and modern humans share the same polymorphisms at the *FOXP2* locus, which differ from those found in all other mammals, suggesting that *Homo sapiens neanderthalensis* most probably had the ability to speak like modern humans. In addition, the reconstructed Neanderthal Y chromosome sequence falls outside the range of variation observed in modern humans, indicating that the *FOXP2* sequences obtained are authentic and confirming that the Neanderthal paternal contribution to the human nuclear genome is low.

During the same period, Lalueza-Fox and colleagues have amplified and sequenced a fragment of the *melanocortin 1 receptor (MC1R)* gene from two Neanderthal remains [84]. The *MC1R* gene regulates pigmentation in humans and other mammals. In particular, *MC1R* variants with a reduced function are associated with pale skin color and red hair in humans of European origin. Lalueza-Fox et al. have shown that both Neanderthal specimens carry a mutation in the *MC1R* gene that is completely absent in the 3700 samples analyzed from modern humans. Furthermore, functional analyses have revealed that this variant reduces *MC1R* activity to a level that alters hair and skin pigmentation in humans. These results suggest that Neanderthals varied in pigmentation levels, potentially on the same scale as observed in modern humans, and, in addition, that defective *MC1R* variants have evolved independently in modern humans and in Neanderthals.

An interesting study by Ludwig et al. [85] based on ancient samples representing different periods between the Late Pleistocene and the Middle Ages and targeting nuclear genes responsible for coat pigmentation, has shed light on the timing and place of horse domestication. This study has shown that the rapid increase in number of coat color patterns is a result of domestication and the great variation in coloration of modern domestic horses is a result of selective breeding by ancient farmers.

The problem of contamination in the classical methodology

Contamination is a serious problem in studies with ancient DNA, even more so when dealing with human

ancient DNA than with animal or plant ancient DNA. Early results from human remains collected in hot climate regions, such as Egypt [2] and Florida [86], are now considered as probably coming from modern human contaminations. For example, a 3.4 kb long nuclear DNA sequence reported as amplified from an Egyptian mummy is an unusually long sequence for DNA extracted from ancient remains. In addition, it is well known that high temperatures such as those occurring in Egypt do not help DNA preservation [87]. Several papers have reported that, despite the use of rigorous protocols [10,88], modern human DNA contamination is prevalent in amplification products from DNA extracts of ancient specimens [13,88-91]. In fact, it has been shown that even with extensive UV and bleach treatments, it is impossible to completely remove modern human DNA from ancient bones and teeth [92]. This is probably due to the porosity of bone and tooth dentine, which are the main entry routes for DNA coming from sweat, skin fragments and exhaled cells. However, when teeth are well preserved and directly removed from the jaw or maxilla, they appear less prone to contamination than bone fragments. Hair is also a more reliable source for studies on human ancient DNA, since it seems to be less susceptible to contamination than bone and teeth, even if its presence is less common in ancient specimens than skeletal remains [93]. In a different study, Salamon et al. [94] has shown that relatively well preserved DNA can be present within clusters of inter-grown bone crystals that are resistant to sodium hypochlorite (NaOCl) disaggregation, a strong oxidant. These authors obtained reproducible authentic DNA sequences from bone crystal aggregates of both modern and ancient animal bones, including humans. Moreover, they claim that the NaOCl treatment minimizes the risk of modern DNA contamination. Indeed, applying this treatment on fossil bones, they have demonstrated the presence of a privileged niche containing DNA in a better state of preservation than DNA extracted from the total bone. Thus, this approach could significantly increase the chances of obtaining authentic ancient DNA sequences, especially from human bones.

For all these reasons, when analyzing ancient human DNA samples with the classical methodology, it is essential to collect skeletal material that has been manipulated only with disposable gloves and face-masks during excavation (Additional file 2), and to select archaeological remains for which the taphonomic history is well-known.

Many excavated archaeological specimens appear to contain DNA from multiple individuals [95,96], which raises the issue of how to authenticate ancient human DNA sequences, when they do not differ from DNA from potential contaminants (modern Human) such as Neanderthals [13] or distinct modern human groups like the Andaman

Islanders [49]. A good example is the analysis of Italian Cro-Magnon specimens [97], which have been submitted to comprehensive authentication protocols [10,88]. However, since the sequences obtained are indistinguishable from those of modern European humans, sample contamination must remain the null hypothesis. Another serious issue is the presence of modern human DNA in samples that were not expected to contain any [13,15,88,94,98-100]. This has been carefully investigated in two studies on Neolithic and Upper Palaeolithic samples, respectively by Sampietro et al. [50] and Caramelli et al. [101]. Sampietro et al. [50] have characterized the mtDNA hypervariable region 1 (HVR1) in 23 Neolithic remains excavated from Granollers (Barcelona, Spain). They compared the sequences obtained with those from the DNA of the six persons involved in the excavation, the washing and subsequent anthropological and genetic studies of the specimens. They found that 17.13 % of the 572 cloned sequences generated from ancient DNA extracted from the teeth of the Neolithic specimens could be unambiguously identified as contaminants derived from the persons involved in the project. Moreover, when checking the cloned contaminant sequences, they observed that the level of damage in the contaminant DNA molecules increased with time and demonstrated that the damage rate of the older contaminant and the endogenous DNA sequences were indistinguishable. Therefore, the commonly used argument that miscoding lesions observed in cloned ancient DNA sequences can be used to support data authenticity is misleading in scenarios where the presence of old contaminant sequences is possible. For this reason, these authors claim that genetic typing of people involved in the manipulation of ancient human specimens is critical to ensure the accuracy and authenticity of findings. In 2008, Caramelli et al. performed the genetic characterization of a Cro-Magnon individual, Paglicci, from layer 23, whose taphonomic history was very well known [101]. Consequently, they were able to monitor all the possible contaminations from persons having manipulated the sample, and demonstrated that the sequence obtained from the Paglicci bones differed from those of all the people that had been in contact with the bones. In addition, since the sequence obtained for the Paglicci mtDNA HVR1 is still common in European humans and since no HVR 1 sequence similar to that of Neanderthals is present in modern Europeans, they conclude that Neanderthals probably did not contribute to the modern human mitochondrial gene pool. In 2010, Krause et al. confirmed this conclusion in a study involving more innovative techniques that will be described in the next sections [12].

Next Generation Sequencing (NGS) methodologies

As described in the above section, in the pre-NGS era, PCR and Sanger sequencing were the main tools available to analyze ancient DNA samples. The

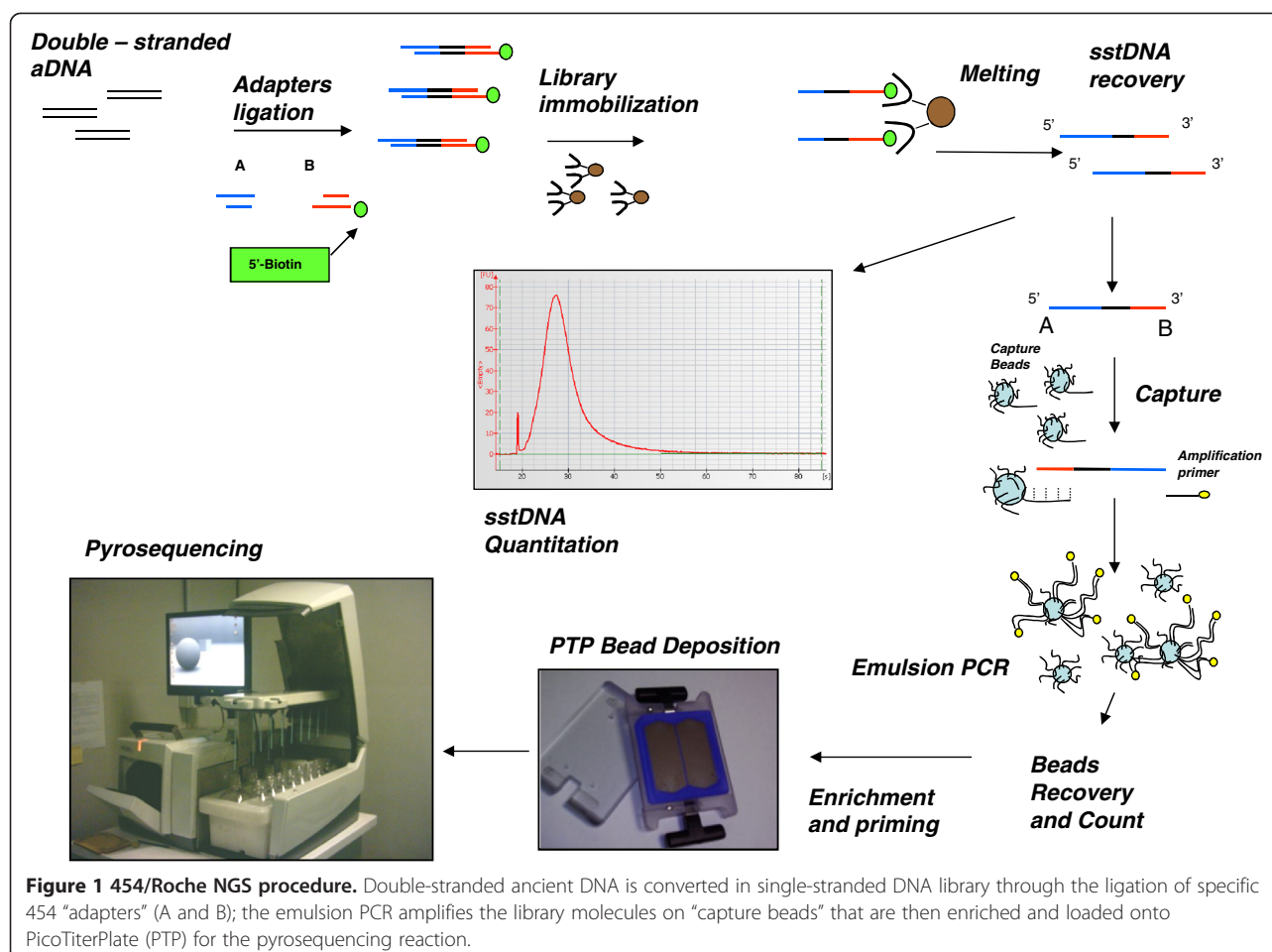
development of miniaturized gel electrophoresis (capillary electrophoresis) and the automation of reactions, gel loading and signal detection allowed the Sanger methodology to become the gold standard for DNA sequencing. Despite these features, Sanger sequencing has a low throughput and consequently is expensive for large-scale sequencing. Moreover, library preparation and amplification, and colony picking have a low efficiency and are time-consuming steps. With ancient DNA, these drawbacks are even more critical and the development of NGS has opened up whole new possibilities and extended the field of applications [102].

NGS methodologies: the background

The new generation of NGS sequencers has made it possible to increase the number of bases sequenced per run with a concomitant decrease in sequencing costs, due to technological improvements. The most important NGS platforms used in the field of ancient DNA analyses are the 454/Roche FLX and the Illumina Genome Analyzer [102-104]. Both technologies share the same rationale for the production of sequences (reads), but differ in the amplification procedure and sequencing chemistry resulting in different throughputs. In spite of the high sensitivity and productivity of NGS sequencers, their signal detection system is not sensitive enough to measure the sequencing signal originating from a single molecule. Detection systems, such as the CCD camera for both 454/Roche and Illumina, can identify a signal only if it is generated by millions of DNA molecules, thus amplification of the sequencing library is necessary. The three key steps for generating reads are: (1) library preparation, (2) library amplification and (3) sequencing. More details about these steps are described and illustrated in Additional file 3 [see Additional file 3] and in Figures 1 and 2.

Library preparation

For both the 454/Roche and Illumina platforms, it is necessary to prepare a library of the ancient DNA fragments ligated at both ends to specific DNA adapters for PCR amplification. After isolating the DNA from an ancient specimen, double strand DNA is polished at the 3' and 5' ends and is converted in blunt end DNA. The polishing step is performed using simultaneously a DNA polymerase and a polynucleotide kinase that catalyzes phosphorylation at the 5' position. Library preparation proceeds with the ligation of the adapters to the polished and phosphorylated fragment ends. Adapters consist of short oligonucleotides of known sequence that will permit the design of complementary primers for library amplification and sequencing.



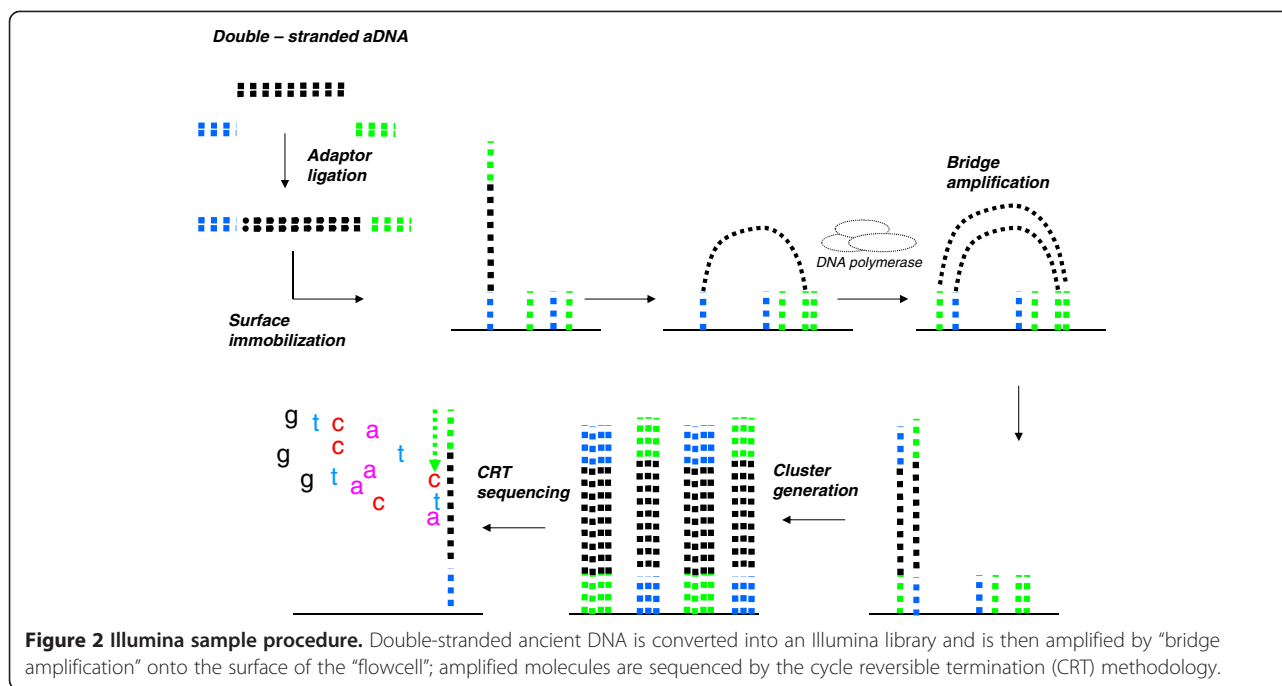
Library amplification

454/Roche and Illumina platforms use different amplification procedures, which are strictly related to their sequencing chemistry. In the case of the 454/Roche platform, the library is amplified in a water-in-oil emulsion PCR (emPCR). Each single DNA molecule of the library is bound to a bead that acts as PCR support, and amplified within an aqueous droplet. In the Illumina system, amplification is carried out by an isothermal bridge amplification process run on a glass slide, termed “flowcell” where molecules are amplified independently, thus generating spatially distinct clones, named “clusters”. While in the Illumina system, the same support i.e. the flowcell is used for both amplification and sequencing, in the 454/Roche system, these two steps are performed on different supports. Further details about amplification procedures are reported in Additional file 3 [see Additional file 3]. Precise library quantification is a critical step for both platforms, to obtain high quality sequences [105]. For the Illumina system, an appropriate number of clusters must be generated during library amplification, while for the 454/Roche system only a precisely quantified library will ensure a correct calibration of DNA/

beads proportions that fosters the binding of a single DNA molecule per bead and the production of an “in-range” emulsion PCR yield.

Sequencing chemistry and throughput

The 454/Roche and Illumina platforms use protocols based on different chemistries and produce different numbers of sequenced bases per run. The strategy to generate reads in the 454/Roche system relies on pyrosequencing, based on a single nucleotide addition (SNA) technique [106,107]. Pyrosequencing consists in an enzymatic cascade that is activated each time a nucleotide is added to the growing DNA chain. Each nucleotide addition generates a spot light that is measured by a CCD camera. The intensity of light emitted by the enzymatic cascade is proportional to the number of added nucleotides and is digitalized and converted in nucleotide sequence. Pyrosequencing reactions take place on a fiber glass support, named “PicoTiterPlate” (PTP), on which DNA beads are loaded after emPCR. The updated version of the 454/Roche GS-FLX platform, Titanium Plus, reaches a read length of 800 bases and each run can generate up to 1 Gigabase of sequence. Conversely, the Illumina platform exploits a sequencing-by-synthesis (SBS) strategy,



in which modified nucleotides, called cyclic reversible terminators (CRT), are inserted during the synthesis reaction. The different fluorescence emissions by each dye-labeled nucleotide, after laser excitation, are detected by a CCD camera and thus each cluster on the flowcell is associated to a raw sequence. The Illumina GAIIx platform has a sequencing yield of about 50 Gigabases per run, while the updated version HiSeq reaches 200 Gigabases per run. The Illumina platform provides sequence reads with a length ranging from 36 to about 150 bases.

Sequencing approaches

In a study on ancient DNA, different types of DNA target can be sequenced depending on the project goal and the available ancient material.

Shotgun sequencing A shotgun sequencing project is performed when isolated DNA is sequenced without any *a priori* selection. The shotgun approach has the capacity of identifying all the known species when total DNA is isolated from bone, teeth or shaft specimen. Given the nature of these sample types, the amount of endogenous target DNA can be very low, because of bacterial or fungal contamination. The shotgun approach is also used for metagenomic studies, when the goal of the sequencing project is to identify all possible known organisms present in an isolated specimen. This approach has been applied by Poinar [108] and Miller [109] to sequence mammoth DNA and they have confirmed the presence of a large amount of exogenous DNA. All the sequences obtained in a shotgun sequencing project are usually identified by

matching to sequence databases, with a "blasting" procedure, that uses *BLAST*. Due to the high statistical power needed to specifically recognize the reads obtained, the shotgun approach requires a high level of sequencing depth, e.g. the total number of bases sequenced has to exceed largely the total length of the target DNA so that each DNA region is sequenced many times. Matching parameters must be carefully chosen to avoid non specific results, as described in the "Data analysis" section.

Considering the cost of this approach, the currently elected sequencing platform is the Illumina Genome Analyzer system, which can generate millions of reads at a lower cost per base compared to the 454/Roche sequencer.

Due to the high level of bacterial contamination in the shotgun approach, a strategy to decrease the amount of microbial DNA sequences has been proposed by Green et al. [11]. It consists in an enzymatic digestion of the sequencing library with a mixture of restriction enzymes capable of degrading DNA fragments with a GC composition similar to that found in bacterial genomes. Green et al. applied this treatment before library amplification and increased by about 5-fold the amount of Neanderthal DNA in sequencing libraries prepared from archeological remains. By applying this method, the same researchers have recently obtained the draft of the entire genome of two extinct hominin groups: Neanderthals and Denisovans [11,110].

Amplicon sequencing PCR is the most efficient procedure to selectively select a target DNA region. PCR includes a series of well-known steps that need to be

optimized for each specific application. The most important step is designing primers, which must take in account the specificity needed to obtain robust results. PCR advantages and disadvantages must be considered before deciding on the rationale of an ancient DNA sequencing project. The most important advantage of a PCR approach is that, via DNA amplification, a high amount of material is obtained i.e. PCR products or amplicons, which are easily converted in an NGS library that can be sequenced. Even if the specificity of PCR is very high, a major disadvantage is the increased rate of misincorporations in the fragment ends, which are important markers for ancient DNA studies, as discussed later in this review.

PCR should be used when the sequencing target is well known and the goal is to detect SNP or small variants used as markers for haplotyping, for instance to identify Neanderthal private nucleotide substitutions. This strategy has been used to identify nucleotide variants in short sequences of specific genes like those determining blood group, taste perception and brain development in Neanderthal samples [111-113]. In a multiplex PCR strategy, different amplicons can be mixed together in a balanced mixture and sequenced at the same time. A multiplex PCR strategy, coupled with NGS sequencing, has made it possible to characterize the entire mitochondrial genome of the Tyrolean Iceman Ötzi, one of the best conserved mummies [114], and to identify the mitochondrial haplotype, K1ö, observed for the first time in this specimen. Another study comparing this sequence with modern human PCR products characterized precisely the nucleotide misincorporations in the mitochondrial DNA of this Tyrolean Iceman Ötzi [115]. Similarly, a PCR multiplexing strategy was used to recover the complete mitochondrial genome sequence of extinct animals, such as cave bears [116] and Auroch (*Bos primigenius*), an ancestor of domestic cattle [117].

Sequence capture Sequence capture is a methodology that uses specifically designed probes to recognize and capture a target DNA, permitting both sample enrichment and recovery of information on DNA misincorporations at the 3' and 5' fragment ends.

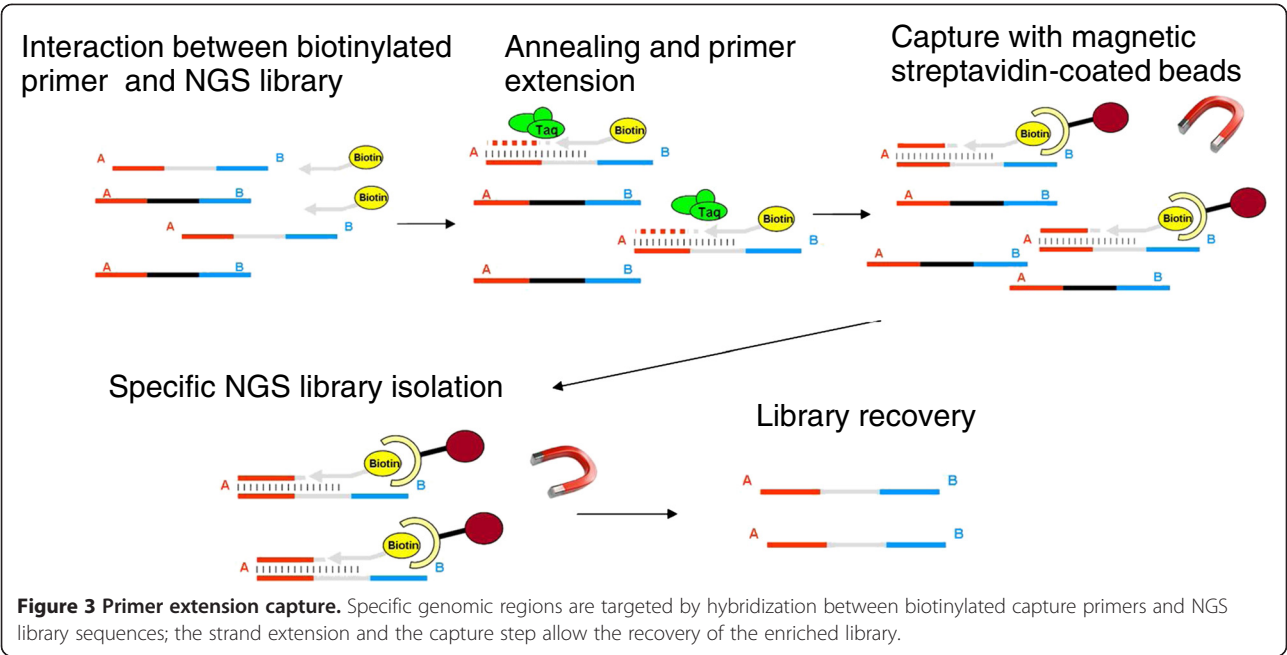
The first capture methodology used in the field of ancient DNA was the primer extension capture (PEC) approach, which uses biotinylated primers specifically designed to identify particular regions and to permit extension until the end of the DNA fragments [118] (Figure 3). Such primers have been exploited by Briggs et al. [118], to capture and sequence the entire Neanderthal mitochondrial genome from whole DNA isolated from an ancient specimen and converted into a labeled NGS library. Once the biotinylated primers bind to the specific target regions contained in the library (e.g. target species mtDNA), an

extension step is performed using a DNA polymerase that synthesizes the complementary strand until the end of the DNA fragment. This extension step increases the specificity of the biotinylated primers. The specifically annealed fragments are captured by streptavidin-coated beads that bind the biotinylated-primer/target-library-fragment complexes. Then, a melting step recovers the captured fraction of the initial library, which represents a small but enriched fraction of the whole DNA isolated from the specimen. This library is finally amplified and sequenced with NGS.

This procedure has a very high specificity and has in fact, improved the capacity to recover complete mtDNA from very complex samples such as Neanderthal and other hominin bone fragments [13,118,119], but synthesizing biotinylated primers for an entire mtDNA sequence is still an expensive step. To decrease the cost of the PEC approach, it is possible to either target only a few short regions or to substitute primer synthesis for probes generated by a cheaper procedure. As shown by Maricic et al. [120], PEC probes can be generated by fragmenting and then biotinylating long products obtained by the amplification of modern mtDNA. This approach makes it possible to obtain PEC probes starting from any target region, whether mitochondrial or nuclear DNA, isolated from human or other animal samples.

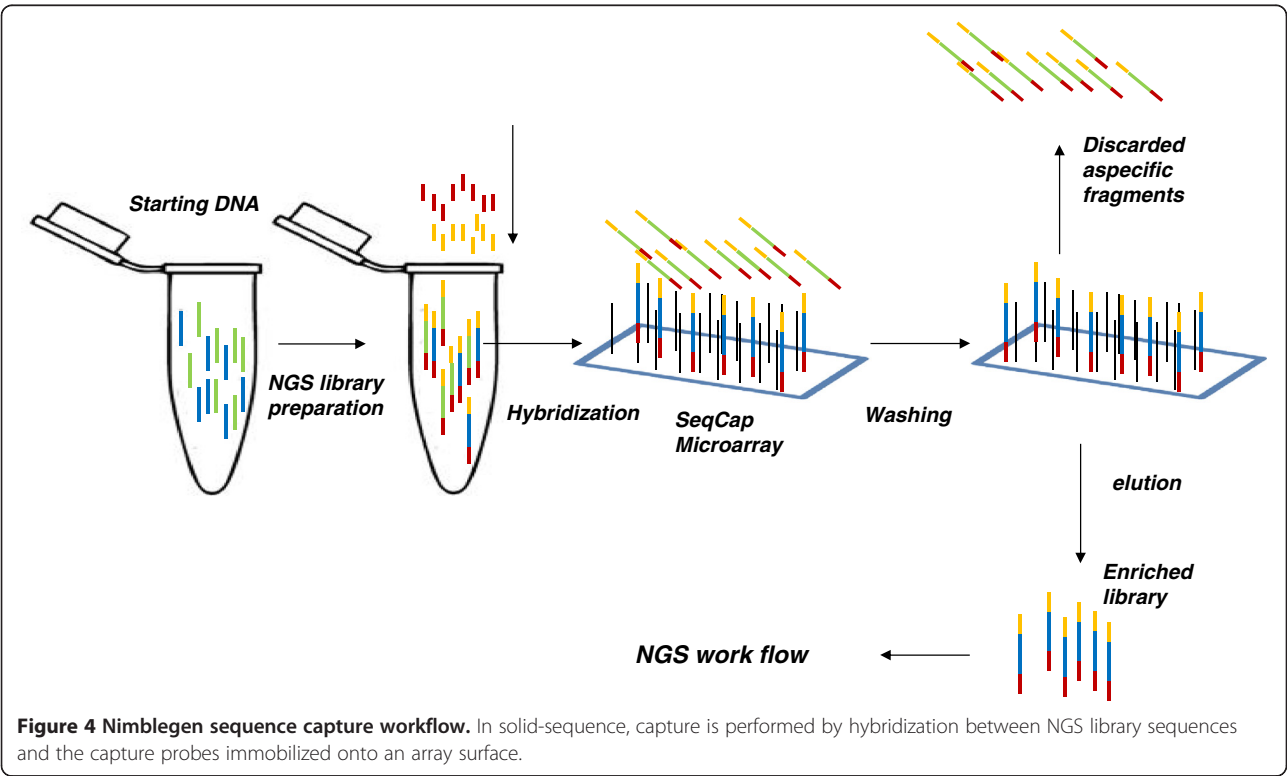
Another capture approach uses biotinylated probes longer than a standard primer, and thus an extension step is not needed and information on misincorporations at the ends of the fragments can be enriched and recovered.

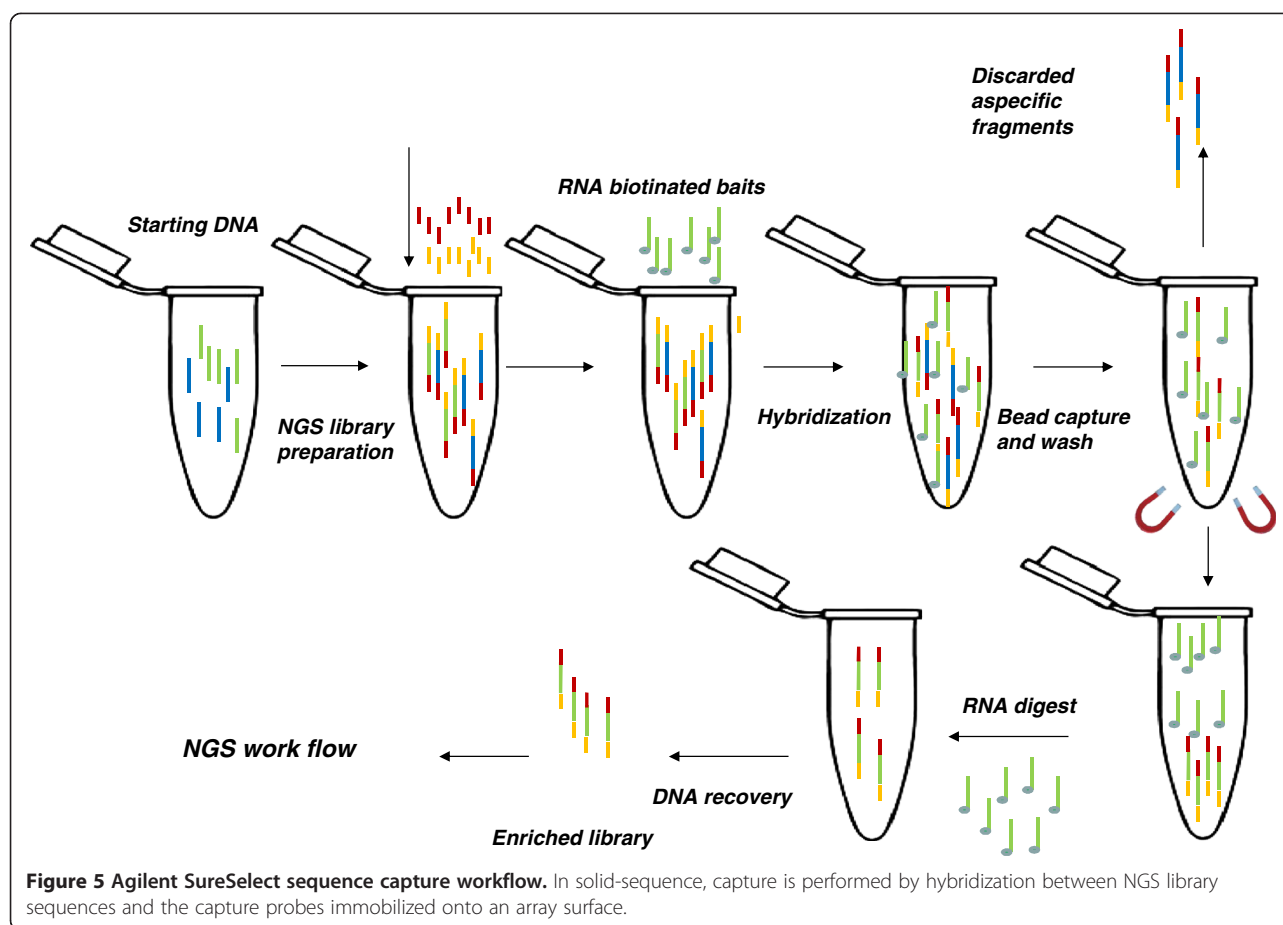
Probe enrichment methodologies are already available for modern human, mouse, dog and cattle DNA re-sequencing projects, which are applied either to study specific genomic regions genetically related to diseases or to analyze a whole exome (i.e. coding exonic DNA) [121,122]. These capture methodologies are available in two forms, either in solid- or in liquid-phase. The solid-phase approach relies on probes immobilized onto an array surface [123], like those used in microarray gene expression studies. Nimblegen (SeqCap arrays) and Agilent (Agilent 244 K microarrays) have developed this kind of approach (Figures 4 and 5 respectively). Burbano and coworkers [124] have used a Nimblegen modified array version for the analysis of ancient DNA. This procedure uses a microarray surface on which probes specifically designed to capture specific genomic regions are immobilized. The NGS library is then hybridized onto the array surface so that the probes capture the targeted DNA regions. After hybridization, non-specific DNA fragments are washed away, while the library DNA fragments complementary to the probes are subsequently recovered by an elution step. These fragments represent a small and selected fraction of



the entire genome. The library enrichment ratio depends on the sizes of the captured genomic regions and of the entire genome. Nimblegen has also developed a commercial capture array, which can capture the whole human exome, with $2.1 \cdot 10^6$ immobilized probes, for a total of $30 \cdot 10^6$ captured bases. Considering the human genome size, this capture array allows a ~ 100 -fold enrichment.

The Agilent in liquid-phase approach i.e. SureSelect target enrichment system, uses biotinylated RNA





probes, named “baits” that are subsequently captured by streptavidin-coated beads [125]. This procedure is similar to that of Nimblegen, but the supporting material for capture consists of beads suspended in a buffer, instead of a solid array. The in-liquid selection has a higher capture yield due to a better interaction between the library and the probes. Another in liquid-phase approach developed by CustomArray uses biotinylated probes synthesized in a customized way using *in situ* synthesis. The DNA probes are then converted into RNA probes and biotinylated to permit enrichment of the target DNA and collection. Finally, the captured enriched library is quantified and sequenced by NGS.

Data analysis

Reference sequence: de-novo or re-sequencing

In general, sequencing projects can be separated into “de-novo” and “re-sequencing” projects, depending on the nature of the sample analyzed and the availability of a reference sequence. In fact, since a reference sequence is required to assemble short reads of damaged DNA, ancient DNA sequencing projects are re-sequencing projects. When a new ancient species is investigated it is

necessary to decide which sequence(s) to use as reference. For anthropological studies, the *Homo sapiens* genome reference sequence is used even when the sequenced ancient DNA originates from a close relative, such as Neanderthal [11] or Denisovan [119]. In these cases, a double mapping onto the human and a close primate genome reference sequences may be required to identify the correct position of the hominin sample in the evolutionary tree. Studies on wild ancestors of modern species generally use the reference sequence of the corresponding modern species [108,117,126].

The large number of on-going sequencing projects and the availability of sequences in public databases, make re-sequencing a widespread approach. However, de-novo sequencing differs from re-sequencing because it requires the generation of an informative and robust sequence with a higher sequencing depth. Thus, the generation of different libraries with different fragment sizes and the assembly of the reads involve the use of specific *ad-hoc* algorithms that can reconstruct a previously unknown sequence.

Mapping of reads Re-sequenced DNA can be analyzed to identify nucleotide variations specific to a given species or to discriminate between phylogenetically-related species.

For example, Neanderthal private polymorphisms in the HVR-1 mitochondrial region are used as markers to discriminate between *Homo sapiens* and Neanderthal DNA [13]. Sequence variations are detected by mapping reads from ancient DNA on a reference sequence using BLAST. Any nucleotide difference between the reference sequence and the ancient DNA reads is a putative private variation. Two important BLAST parameters must be considered to obtain robust results: the gap opening penalty (g) and the e -value (e). When analyzing ancient DNA sequences, it is necessary to avoid setting too stringent matching parameters, to increase the probability of having an informative result. Moreover, sequencing errors must be taken into account when choosing matching parameters. Illumina and 454/Roche have different patterns of sequencing errors. On the one hand, Illumina sequencing is associated with a high error in base calling, which leads to a high rate of false base substitutions. These errors can be detected and corrected by sequencing at a sufficient depth (20X or more). On the other hand, the 454/Roche pyrosequencing reaction cascade accumulates insertion or deletions (INDELs) within homopolymeric regions resulting in a sequence of low quality in these regions. Thus, putative variants detected in these regions will require further validation. Given the sequencing error rates of these two NGS platforms, many sequencing projects combine the two technologies to compensate for their respective limits.

Considering all the difficulties associated with sequencing ancient DNA, such as nucleotide misincorporations and NGS error rates, matching parameters G and e -values, for reads mapping should be carefully chosen. For these reasons, the parameter G that is related to the space introduced during the alignment between two sequences and represents the cost to open a gap is usually set to low values. Moreover, the e -value that is a parameter related to the reliability of the similarity between two sequences is expected to be very low for specific matches. Matching results must be filtered to discard data with non-specific matches.

Mind the gap The origin of the reads is variable since they can originate from both ancient and modern DNA fragments present in the analyzed sample. Currently, different criteria are used to recognize and discriminate ancient from modern DNA. Ancient DNA fragments are usually short, less than 120 bp long, but this is not sufficient to distinguish between ancient and modern DNA. Another important feature of ancient DNA is the presence of miscoding lesions at the 5' and 3' ends of the fragments [127], where a higher frequency of base substitutions from C to T and from G to A is observed compared to that in modern DNA (approximately 35 % in Pleistocene ancient human samples). Thus, a reduced fragment length associated with the occurrence of terminal miscoding lesions is the main

feature used to distinguish ancient from modern DNA. If a high sequencing depth is available, DNA damage rate can be calculated with the "mapDamage" software [128]. Contamination from modern DNA requires special attention when hominin specimens are investigated. Even if the sample's taphonomic history is well known, putative contaminations from modern human DNA can reduce the number of informative reads. The shotgun approach can discriminate between target and contaminant reads, but the cost to obtain robust results can be very high. Using a selection procedure such as sequence capture can reduce the spectrum of target sequences, while keeping information on miscoding lesions in fragments of ancient DNA.

Applications of NGS technologies in the field of ancient DNA

From the mammoth to the Neanderthal genome project

The first publication on ancient DNA sequencing using an NGS platform was entitled "Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA", published by Poinar et al. in *Science* at the end of 2005 [108]. Poinar and colleagues performed a metagenomic study on a mammoth bone sample using the first version of the 454 sequencer and produced 302 692 sequence reads with a mean length of 95 bp, for a total of nearly 30.10^6 bases. After this milestone publication, many other sequencing projects of ancient DNA have been carried out based on high-throughput next-generation sequencing.

In May 2006, the first nuclear DNA sequences from a Neanderthal (*Homo neanderthalensis*) were reported by Pääbo [129], as part of the Neanderthal Genome project that had started about two years before. Within this project, Pääbo and coworkers probed 60 Neanderthal specimens from museums to investigate the degree of DNA preservation after thousands of years. Two of the specimens analyzed provided interesting results and Pääbo's team reported at the Biology of Genomes meeting at New York's Cold Spring Harbor Laboratory that they had succeeded in sequencing about 10^6 bp of nuclear DNA (about 0.03 % of the genome) from a 45 000-year-old male from the Vindija cave, Croatia using pyrosequencing. Another report by James Noonan at the Cold Spring Harbor meeting in the same year, presented a preliminary analysis of 75 000 of the 10^6 bp sequenced so far. These two papers [129,130] on Neanderthal nuclear DNA sequences from fossil bones promise to answer questions on the relationship between extinct species, such as Neanderthals and present-day humans. Typically, one of these questions is whether the Neanderthals are the direct ancestors of modern-day humans or an evolutionary side branch that eventually died out. Although these two studies were the first to investigate *Homo* ancient DNA on a large scale, they came to very different conclusions regarding the ancestral role of Neanderthals. A year later, Wall and Kim [131] reanalyzed the

data from these two original studies and found that their results contradict each other. This implies that the data from at least one of the studies are probably incorrect, due to contamination by modern human DNA, which strongly compromises the final findings [131].

Recently, Green et al. [11] sequenced a large part of the Neanderthal nuclear genome using NGS. Starting from about 400 mg of bone powder, they generated 5.3 Gigabases of Neanderthal DNA sequences with a 1.3X genome coverage. DNA sequences matching at a significantly higher rate to primate genomes than to any other organism were further analyzed. These results revealed that Neanderthals and present-day humans shared a common ancestor about 800 000 years ago, which diverged between 270 000 and 440 000 years ago. This study also showed that Neanderthal DNA shares more genetic variants with present-day humans from Eurasia than from sub-Saharan Africa i.e. on average 2.5 % (range 1–4 %) of the genome of people outside Africa derive from Neanderthals. In addition, the authors provided a list of genomic regions and candidate genes that could have been under early positive selection pressure in modern human history, for example, those involved in cognitive abilities and cranial morphology.

Denisova man Along these same lines, another important study has shed new light on the evolutionary history of the *Homo* genus. In 2010, Krause et al. have described the distal manual phalanx of a juvenile hominin excavated from the Denisova cave (Altai Mountains, southern Siberia) [119]. The phalanx was found in layer 11, which has been dated 50 000 to 30 000 years ago. This layer contained micro blades and body ornaments of polished stone, typical of the 'Upper Palaeolithic industry' and generally thought to be associated with modern humans, and also stone tools that are more characteristic of the earlier Middle Paleolithic age, such as side scrapers and Levallois blanks. A sequence capture approach [118,119] combined with high-throughput sequencing [104,119] was used to characterize the complete mtDNA genome from this specimen.

To clarify the relationship between the Denisova individual and other hominin groups, Reich et al. have sequenced the nuclear genome of the Denisova specimen and analyzed its genomic relationship with Neanderthal and present-day human genomes [110]. Two independent sequencing libraries were generated from the ancient DNA extracted from this specimen, using a modified Illumina protocol. The results show that the Denisova individual belongs to a hominin group, named the "Denisovans", which share a more recent common ancestor with Neanderthal but have a population history distinct from both Neanderthals and modern humans. The major outcome of this study is

that this is the first time that a new population, the "Denisovans" is characterized starting from DNA. In fact, at present, except for the phalanx and a tooth retrieved during the archaeological excavation, no other physical and morphological feature of this population is known. Additional important issues have been addressed in a study on archaic hominins taking into consideration the important argument of gene flow between modern and extinct humans, as reported by Lalueza-Fox et al. [132].

Modern human evolution With the advent of the NGS technology and the possibility to reconstruct the genome of extinct hominins, new perspectives to study human evolution have opened up. Malmström et al. have studied the dynamics of a Scandinavia population during the Neolithic era by sequencing DNA from ancient Scandinavian human remains [133]. Gilbert et al. have sequenced the mtDNA of a Paleo-Eskimo human starting from a 3400 to 4500-year-old frozen hair, thus documenting human expansion into the New World's northern extremes [134]. The same DNA source (i.e. hair shaft) had already been used by Gilbert et al. [135] to sequence the mitochondrial genome of a Siberian mammoth (*Mammuthus primigenius*) using NGS. Interesting results have been reported by Rasmussen et al. who sequenced almost 80 % of the genome of a 4000-year-old human genome from Greenland [136]. Besides from showing that the Saqqaq population had no or little European contribution in their genome, this sequencing project identified specific phenotypic traits starting from genomic information, providing important information about an extinct culture. Although these genomic data are obtained from a single individual, a demographic history could be reconstructed.

Domesticated and extinct species Using NGS, it is possible to detect traces of the intermingled genetic histories of the humans involved in the Neolithic processes that led to the domestication and diffusion of a species, as shown by the sequencing of the mitochondrial genome of *Bos primigenius* [61,117].

Using the Roche/454 platform, Ramírez et al. [26] have sequenced DNA isolated from a 6000-year-old bone sample from an extinct caprine species (*Myotragus balearicus*) from the Balearic Islands and thus shown that NGS applied to DNA from extinct animals can lead to a better understanding of adaptation.

Other extinct animals, such as the New Zealand moa have been characterized by Allentoft et al. [137]. They have demonstrated the feasibility of studying ancient biodiversity and extinction processes using NGS. In addition, Miller et al. [138] have analyzed the Tasmanian tiger (*Thylacinus cynocephalus*), extinct since 1936, by sequencing DNA isolated from a museum specimen collection.

Third Generation Sequencers (TGS) and ancient DNA: from the past to the future

With the development of new sequencing chemistries and innovative signal detection methodologies, a new generation of sequencers, named Third Generation Sequencers (TGS) have recently become available (Table 2). These novel platforms differ in sample preparations and sequencing chemistries, but all share the particular and innovative feature of direct sequencing, allowing single molecule analysis, hence bypassing PCR amplification. For studies on ancient DNA, the possibility of using very small amounts of starting DNA field is highly appreciated, since it improves detection of DNA damage and identification of contaminants and exogenous sequences. Recently, Orlando and colleagues [139] have applied this new technology to perform single-molecule sequencing of DNA isolated from a Pleistocene horse bone. They used the total DNA isolated from the specimen, without any kind of sequence enrichment (i.e. sequence capture), or PCR amplification, thus allowing an assessment of postmortem DNA damage. In spite of a very high number of sequenced bases, the level of sequence depth was too low (1X on the mitochondrial genome) to proceed further with the analysis Table 2.

Conclusions

The second generation sequencers, such as Roche/454 and Illumina presented in this review but also SOLiD have opened new possibilities for ancient DNA sequencing by providing a very high throughput, which could not be reached previously with traditional Sanger-based platforms. In the last few years, these technologies have made it possible to characterize genomes, metagenomes and transcripts, thus rapidly increasing sequence information available for different organisms and applications. The three most important NGS technologies have different specific features positioning each one in a given niche market. Briefly, the Roche/454 technology provides long sequence reads, today up to 800 bases, thus greatly simplifying the genome assembly. The Illumina and SOLiD technologies, thanks to their high throughput, have drastically reduced the cost per sequenced base even if the reads produced do not exceed 100 bases. Although, to date, the SOLiD technology has not been used in studies on ancient DNA, it

offers an important asset i.e. high accuracy in base calling due to its two-base encoding sequencing strategy. The throughput of the SOLiD platform, based on a sequencing by ligation (SBL) chemistry is comparable to that obtained by Illumina, but because a reference sequence is required to translate its so-called “color space” into a nucleotide sequence, its application in the field of ancient DNA has not been considered up till now.

The high sensitivity and resolution power offered by the NGS technologies have considerably increased available data on ancient DNA thus greatly improving our knowledge on many interesting issues of human evolution: the history of extinct species, adaptation and domestication processes. New hypotheses and theories have been developed with the most important achievement being the robustness and reliability of sequencing data. Moreover, identification, elimination and estimation of contaminant sequences in ancient DNA samples are issues that need to be discussed and resolved. Developing new molecular biology procedures and innovative *ad-hoc* bioinformatics approaches must be accounted for when planning a sequencing project on ancient DNA. Each feature of an ancient sample needs to be carefully evaluated to develop a specifically focused experimental design in order to optimize the sequencing power of either the NGS or TGS technology used. For this reason, when NGS is used to sequence ancient DNA, we propose to include an additional “golden criteria” to the nine summarized by Cooper and Poinar in 2000 [10]. It concerns the labeling of the ancient DNA sequencing library using synthetic commercial or custom-made nucleotides “TAG”. These “labels” could be project- or sample-specific in the case of shotgun or capture sequencing, to enable estimation of putative contamination of the sample after DNA isolation.

Currently, enrichment protocols of DNA samples and specific bioinformatics pipelines have been successfully developed to increase data reliability. However, other steps in the procedures and analyses need to be optimized to generate a robust ancient DNA data set. NGS technologies and the development of new molecular biology strategies, such as PEC or capture “baits”, have also accelerated the change in the focus of research from the frequently investigated mtDNA (reviewed by Ho and Gilbert [140]) to the more informative nuclear DNA. Major insights have been gained from the multidisciplinary nature of the projects on ancient DNA, which exploit data combined from different technologies capable of generating genome-wide data [141]. Moreover, all the data, which have been used to study complex processes such as human evolution, domestication or demographic events, need to be confirmed by the analysis of many more samples, even if the availability of ancient well-preserved specimens is always a critical and intrinsic feature of palaeogenomic projects.

Table 2 Third Generation Sequencers (TGS): an overview

Company name	TGS sequencing chemistry
Helicos Genetic Analysis Platform	Virtual Terminator nucleotides
Pacific Biosciences	Anchored DNA polymerase + Zero-mode waveguide (ZMW)
VisiGen Biotechnologies	Modified DNA polymerase + Fluorescence Resonance Energy Transfer (FRET)

Company name and corresponding technology are listed.

Additional files

Additional file 1: The « Golden criteria » by Cooper and Poinar 2000 [10].

Additional file 2: How samples should be collected.

Additional file 3: NGS library preparation and sequencing chemistries. The material provided describes library preparation for the 454/Roche and Illumina platforms and their sequencing chemistries.

Competing interests

The authors declare that they have no competing interest.

Acknowledgements

This study was supported by MIUR grant #2008TE2B8S_002 and by Progetti di ricerca fondamentale o di base Sardinia Region #CRP2_439 assigned to DC, FIRB grant #RBF08U07M "Futuro in Ricerca" assigned to ER. We are grateful to Roscoe Stanyon for revising manuscript.

Author details

¹Institute for Biomedical Technologies, National Research Council, Via F.lli Cervi 93, Segrate, Milan 20090, Italy. ²Department of Evolutionary Biology, Laboratory of Anthropology, Molecular Anthropology/Paleogenetics Unit, University of Florence, Via del Proconsolo 12, Florence 50122, Italy.

Authors' contributions

DC conceived the project. ER, ML, EG, GDB and DC wrote the manuscript. All authors read and approved the final manuscript.

Received: 6 September 2011 Accepted: 14 June 2012

Published: 14 June 2012

References

- Higuchi R, Bowman B, Freiburger M, Ryder OA, Wilson AC: **DNA sequences from the quagga, an extinct member of the horse family.** *Nature* 1984, **312**:282–284.
- Pääbo S: **Molecular cloning of ancient Egyptian mummy DNA.** *Nature* 1985, **314**:644–645.
- Pääbo S: **Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification.** *Proc Natl Acad Sci USA* 1989, **86**:1939–1943.
- Pääbo S, Higuchi RG, Wilson AC: **Ancient DNA and the polymerase chain reaction.** *J Biol Chem* 1989, **264**:9709–9712.
- Thomas RH, Schaffner W, Wilson AC, Pääbo S: **DNA phylogeny of the extinct marsupial wolf.** *Nature* 1989, **340**:465–467.
- Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G: **Chloroplast DNA from a Miocene Magnolia species.** *Nature* 1990, **344**:656–658.
- Soltis PS, Soltis DE, Smiley CJ: **An rbcL sequence from a Miocene Taxodium (bald cypress).** *Proc Natl Acad Sci USA* 1992, **89**:449–451.
- Kim S, Soltis DE, Soltis PS, Sue Y: **DNA sequences from Miocene fossils: an ndhF sequence of Magnolia latahensis (Magnoliaceae) and an rbcL sequence of Persea pseudocarolinensis (Lauraceae).** *Am J Bot* 2004, **91**:615–620.
- Woodward SR, Weyand NJ, Bunell M: **DNA sequence from Cretaceous period bone fragments.** *Science* 1994, **266**:1229–1232.
- Cooper A, Poinar HN: **Ancient DNA: Do it right or not at all.** *Science* 2000, **289**:1139.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S: **A draft sequence of the Neanderthal genome.** *Science* 2010, **328**:710–722.
- Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, Pääbo S: **A complete mtDNA genome of an early modern human from Kostenki, Russia.** *Curr Biol* 2010, **20**:231–236.
- Krings M, Stone A, Schmitz RW, Krainitzk H, Stoneking M, Pääbo S: **Neandertal DNA sequences and the origin of modern humans.** *Cell* 1997, **90**:19–30.
- Krings M, Geisert H, Schmitz RW, Krainitzk H, Pääbo S: **DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen.** *Proc Natl Acad Sci USA* 1999, **96**:5581–5585.
- Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martínez-Maza C, Bastir M, Fortea J, de la Rasilla M, Bertranpetit J, Rosas A: **Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian peninsula.** *Mol Biol Evol* 2005, **22**:1077–1081.
- Caramelli D, Lalueza-Fox C, Condemi S, Longo L, Milani L, Manfredini A, DeSaint Pierre M, Adoni F, Lari M, Giunti P, Ricci S, Casoli A, Calafell F, Mallegni F, Bertranpetit J, Stanyon R, Bertorelle G, Barbujani G: **A highly divergent DNA sequence in a Neandertal individual from Italy.** *Curr Biol* 2006, **16**:R630–R632.
- Lalueza-Fox C, Krause J, Caramelli D, Catalano G, Milani L, Sampietro ML, Calafell F, Martínez-Maza C, Bastir M, García-Tabernero A, De la Rasilla M, Fortea J, Pääbo S, Bertranpetit J, Rosas A: **Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals.** *Curr Biol* 2006, **16**:R629–R630.
- Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, Hänni C: **Revisiting Neandertal diversity with a 100 000 year old mtDNA sequence.** *Curr Biol* 2006, **16**:R400–R402.
- Krause J, Orlando L, Serre D, Viola B, Prüfer K, Richards MP, Hublin JJ, Hänni C, Derevianko AP, Pääbo S: **Neanderthals in central Asia and Siberia.** *Nature* 2007, **449**:902–904.
- Krajewski C, Buckley L, Westerman M: **DNA phylogeny of the marsupial wolf resolved.** *Proc R Soc London, Ser B* 1997, **264**:911–917.
- Cooper A, Mourer-Chauvire C, Chambers GK, von Haeseler A, Wilson AC, Pääbo S: **Independent origins of New Zealand moas and kiwis.** *Proc Natl Acad Sci USA* 1992, **89**:8741–8744.
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R: **Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution.** *Nature* 2010, **409**:704–707.
- Greenwood AD, Castresana J, Feldmaier Fuchs G, Pääbo S: **A molecular phylogeny of two extinct sloths.** *Mol Phylogenet Evol* 2001, **18**:94–103.
- Paxinos EE, James HF, Olson SL, Sorenson MD, Jackson J, Fleischer RC: **MtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (Branta canadensis).** *Proc Natl Acad Sci USA* 2002, **99**:1399–1404.
- Hänni C, Laudet V, Stehelin D, Taberlet P: **Tracking the origins of the cave bear (Ursus spelaeus) by mitochondrial DNA sequencing.** *Proc Natl Acad Sci USA* 1994, **91**:12336–12340.
- Ramírez O, Gigli E, Bover P, Alcover JA, Bertranpetit J, Castresana J, Lalueza-Fox C: **Paleogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine.** *PLoS One* 2009, **4**:e5670.
- Orlando L, Calvignac S, Schönebelen C, Douady CJ, Godfrey LR, Hänni C: **DNA from extinct giant lemurs links archaiolemurids to extant indriids.** *BMC Evol Biol* 2008, **8**:121.
- Driscoll CA, Yamaguchi N, Bar-Gal GK, Roca AL, Luo S, Macdonald DW, O'Brien SJ: **Mitochondrial phylogeography illuminates the origin of the extinct Caspian tiger and its relationship to the amur tiger.** *PLoS One* 2009, **4**:e4125.
- Thomas WK, Pääbo S, Villablanca FX, Wilson AC: **Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens.** *J Mol Evol* 1990, **31**:101–112.
- Pergams ORW, Barnes WM, Nyberg D: **Rapid change in mouse mitochondrial DNA.** *Nature* 2003, **423**:397.
- Hardy C, Callou C, Vigne JD, Casane D, Dennebouy N, Mounolou JC, Monnerot M: **Rabbit mitochondrial DNA diversity from prehistoric to modern times.** *J Mol Evol* 1995, **40**:227–237.
- Hadly EA, Kohn MH, Leonard JA, Wayne RK: **A genetic record of population isolation in pocket gophers during Holocene climatic change.** *Proc Natl Acad Sci USA* 1998, **95**:6893–6896.
- Wisely SM, Buskirk SW, Fleming MA, McDonald DB, Ostrander EA: **Genetic diversity and fitness in black-footed ferrets before and during a bottleneck.** *J Hered* 2002, **93**:231–237.
- Larson S, Jameson R, Etnier M, Fleming M, Bentzen P: **Loss of genetic diversity in sea otters (Enhydra lutris) associated with the fur trade of the 18th and 19th centuries.** *Mol Ecol* 2002, **11**:1899–1903.
- Miller CR, Waits LP: **The history of effective population size and genetic diversity in the Yellowstone grizzly (Ursus arctos): implications for conservation.** *Proc Natl Acad Sci USA* 2003, **100**:4334–4339.

36. Hale ML, Lurz PW, Shirley MD, Rushton S, Fuller RM, Wolff K: **Impact of landscape management on the genetic structure of red squirrel populations.** *Science* 2001, **293**:2246–2248.
37. Verginelli F, Capelli C, Coia V, Musiani M, Falchetti M, Ottini L, Palmirotta R, Tagliacozzo A, De Grossi Mazzorin I, Mariani-Costantini R: **Mitochondrial DNA from prehistoric canids highlights relationships between dogs and South-East European wolves.** *Mol Biol Evol* 2005, **22**:2541–2551.
38. Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C: **Rates of evolution in ancient DNA from Adelie penguins.** *Science* 2002, **295**:2270–2273.
39. Dalén L, Nyström V, Valdiosera C, Germonpré M, Sablin M, Turner E, Angerbjörn A, Arsuaga JL: **Götherström: Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox.** *Proc Natl Acad Sci USA* 2007, **104**:6726.
40. Lawrence HA, Scofield RP, Crockett DE, Millar CD, Lambert DM: **Ancient genetic variation in one world's rarest seabirds.** *Heredity* 2008, **101**:543–547.
41. Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Gröcke DR, Matheus P, Zazula G, Guthrie D, Froese D, Buigues B, de Marliave C, Flemming C, Poinar D, Fisher D, Southon J, Tikhonov AN, MacPhee RD, Poinar HN: **Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths.** *Curr Biol* 2008, **18**:1320–1326.
42. Krause J, Unger T, Nögon A, Malaspinas AS, Kolokotronis SO, Stiller M, Soibelman L, Spriggs H, Dear PH, Briggs AW, Bray SC, O'Brien SJ, Rabeder G, Matheus P, Cooper A, Slatkin M, Pääbo S, Hofreiter M: **Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary.** *BMC Evol Biol* 2008, **8**:220.
43. Leonard JA, Rohland N, Glaberman S, Fleischer RC, Caccone A, Hofreiter M: **A rapid loss of stripes: the evolutionary history of the extinct quagga.** *Biol Lett* 2005, **22**:291–295.
44. Steeves TE, Holdaway RN, Hale ML, McIlroy E, McAllan IA, Christian M, Hauber ME, Bunce M: **Merging ancient and modern DNA: extinct seabird taxon rediscovered in the North Tasman Sea.** *Biol Lett* 2010, **6**:94–97.
45. Seabrook-Davison M, Huynen L, Lambert DM, Brunton DH: **Ancient DNA resolves identity and phylogeny of New Zealand's extinct and living quail (*Coturnix* sp.).** *PLoS One* 2009, **4**:e6400.
46. Ramakrishnan U, Hadly EA: **Using phylogenomics to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies.** *Mol Ecol* 2009, **18**:1310–1330.
47. Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A: **Dynamics of Pleistocene population extinctions in Beringian brown bears.** *Science* 2002, **295**:2267–2270.
48. Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, Cappellini E, Moggi-Cecchi J, Chiarelli B, Castri L, Casoli A, Mallegni F, Lalueza-Fox C, Barbujani G: **The Etruscans: a population-genetic study.** *Am J Hum Genet* 2004, **74**:694–704.
49. Endicott P, Gilbert MTP, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ: **The genetic origins of the Andaman Islanders.** *Am J Hum Genet* 2003, **72**:178–184.
50. Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, Lari M, Agusti B, Bertranpetit J, Lalueza-Fox C: **The genetics of the pre-Roman Iberian Peninsula: a mtDNA study of ancient Iberians.** *Ann Hum Genet* 2005, **69**:535–548.
51. Caramelli D, Vernesi C, Sanna S, Sampietro L, Lari M, Castri L, Vona G, Floris R, Francalacci P, Tykot R, Casoli A, Bertranpetit J, Lalueza-Fox C, Bertorelle G, Barbujani G: **Genetic variation in prehistoric Sardinia.** *Hum Genet* 2007, **122**:327–336.
52. Guimaraes S, Ghirotto S, Benazzo A, Milani L, Lari M, Pilli E, Pecchioli E, Mallegni F, Lippi B, Bertoldi F, Gelichi S, Casoli A, Belle EM, Caramelli D, Barbujani G: **Genealogical discontinuities among Etruscan, Medieval and contemporary Tuscans.** *Mol Biol Evol* 2009, **26**:2157–2166.
53. Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, Burger J: **Ancient DNA from the first European farmer in 7500-year-old Neolithic sites.** *Science* 2005, **310**:1016–1018.
54. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J: **Genetic discontinuity between local hunter-gatherers and central Europe's first farmers.** *Science* 2009, **326**:137–140.
55. Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CS, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt KW, Cooper A: **Members of the Genographic Consortium: Ancient DNA from European early neolithic farmers reveals their near eastern affinities.** *PLoS Biol* 2010, **8**:e1000536.
56. Bailey JF, Richards MB, Macaulay VA, Colson IB, James IT, Bradley DG, Hedges RE, Sykes BC: **Ancient DNA suggest a recent expansion of European cattle from a diverse wild progenitor species.** *Proc Biol Sci* 1996, **263**:1467–1473.
57. Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC, Bradley DG: **Genetic evidence for Near-Eastern origins of European cattle.** *Nature* 2001, **410**:1088–1091.
58. Anderung C, Bouwman A, Persson P, Carretero JM, Ortega AI, Elburg R, Smith C, Arsuaga JL, Ellegren H, Götherström A: **Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle.** *Proc Natl Acad Sci USA* 2005, **102**:8431–8435.
59. Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, Casoli A, Goyache F, Royo LJ, Conti S, Lari M, Martini A, Ouragh L, Magid A, Atash A, Zsolnai A, Boscato P, Triantaphyllidis C, Ploumi K, Sineo L, Mallegni F, Taberlet P, Erhardt G, Sampietro L, Bertranpetit J, Barbujani G, Luikart G, Bertorelle G: **The origin of European cattle: Evidence from modern and ancient DNA.** *Proc Natl Acad Sci USA* 2006, **103**:8113–8118.
60. Bollongino R, Edwards CJ, Alt KW, Burger J, Bradley DG: **Early history of European domestic cattle as revealed by ancient DNA.** *Biol Lett* 2006, **2**:155–159.
61. Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, Vigne JD, Baird JF, Larson G, Ho SY, Heupink TH, Shapiro B, Freeman AR, Thomas MG, Arbogast RM, Arndt B, Bartosiewicz L, Benecke N, Budja M, Chaix L, Choyke AM, Coquegniot E, Döhle HJ, Göldner H, Hartz S, Helmer D, Herzig B, Hongo H, Mashkour M, Ozdogan M, Pucher E, Roth G, Schade-Lindig S, Schmölcke U, Schulting RJ, Stephan E, Uerpmann HP, Vörös I, Voytek B, Bradley DG, Burger J: **Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs.** *Proc Biol Sci* 2007, **274**:1377–1385.
62. Mona S, Catalano G, Lari M, Larson G, Boscato P, Casoli A, Sineo L, Di Patti C, Pecchioli E, Caramelli D, Bertorelle G: **Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion.** *BMC Evol Biol* 2010, **10**:83.
63. Achilli A, Olivieri A, Pellicchia M, Uboldi C, Colli L, Al-Zahery N, Accetturo M, Pala M, Kashani BH, Perego UA, Battaglia V, Fornarino S, Kalamati J, Houshmand M, Negrini R, Semino O, Richards M, Macaulay V, Ferretti L, Bandelt HJ, Ajmone-Marsan P, Torroni A: **Mitochondrial genomes of extinct aurochs survive in domestic cattle.** *Curr Biol* 2008, **18**:R157–R158.
64. Achilli A, Bonfiglio S, Olivieri A, Malusà A, Pala M, Kashani BH, Perego UA, Ajmone-Marsan P, Liotta L, Semino O, Bandelt HJ, Ferretti L, Torroni A: **The multifaceted origin of taurine cattle reflected by the mitochondrial genome.** *PLoS One* 2009, **4**:e5753.
65. Bonfiglio S, Achilli A, Olivieri A, Negrini R, Colli L, Liotta L, Ajmone-Marsan P, Torroni A, Ferretti L: **The enigmatic origin of bovine mtDNA haplogroup R: sporadic interbreeding or an independent event of Bos primigenius domestication in Italy?** *PLoS One* 2010, **5**:e15760.
66. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A: **Worldwide phylogeography of wild boar reveals multiple centers of pig domestication.** *Science* 2005, **307**:1618–1621.
67. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne JD, Edwards CJ, Schlumbaum A, Dinu A, Balacescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG, Cooper A: **Ancient DNA pig domestication and the spread of the Neolithic into Europe.** *Proc Natl Acad Sci USA* 2007, **104**:15276–15281.
68. Gongora J, Rawlence NJ, Mobegi VA, Jianlin H, Alcalde JA, Matus JT, Hanotte O, Moran C, Austin JJ, Ulm S, Anderson AJ, Larson G, Cooper A: **Indo-European and Asian origins for Chilean and Pacific chickens revealed by mt DNA.** *Proc Natl Acad Sci USA* 2008, **105**:10308–10313. Erratum in: *Proc Natl Acad Sci USA* 2008, **105**:14234.
69. Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, Reissmann M, Ludwig A: **Origin and history of mitochondrial DNA lineages in domestic horses.** *PLoS One* 2010, **5**:e15311.
70. Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Pääbo S, Hofreiter M: **Multiplex amplification of the**

- mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 2006, **439**:724–727.
71. Hofreiter M, Jaenicke V, Serre D: **Haeseler Av A, Pääbo S: DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA.** *Nucl Acids Res* 2001, **29**:4793–4799.
 72. Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A: **Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions.** *Nucl Acids Res* 2007, **35**:5717–5728.
 73. Vestheim H, Jarman SN: **Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs.** *Front Zool* 2008, **5**:12.
 74. Gigli E, Rasmussen M, Civit S, Rosas A, de la Rasilla M, Fortea J, Gilbert MTP, Willerslev E, Lalueza-Fox C: **An improved PCR method for endogenous DNA retrieval in contaminated Neandertal samples based on the use of blocking primers.** *J Archaeol Sci* 2009, **36**:2676–2679.
 75. Greenwood AD, Capelli C, Plessner G, Pääbo S: **Nuclear DNA sequences from Late Pleistocene megafauna.** *Mol Biol Evol* 1999, **16**:1466–1473.
 76. Poinar H, Kuch M, McDonald G, Martin P, Pääbo S: **Nuclear gene sequences from a Late Pleistocene sloth coprolite.** *Curr Biol* 2003, **12**:1150–1152.
 77. Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MT, Cooper A, Doebley J, Pääbo S: **Early allelic selection in maize as revealed by ancient DNA.** *Science* 2003, **302**:1206–1208.
 78. Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, Drummond A, Cooper A: **Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*.** *Nature* 2003, **425**:172–175.
 79. Huynen L, Millar CD, Scofield RP, Lambert DM: **Nuclear DNA sequences detect species limits in ancient moa.** *Nature* 2003, **425**:175–178.
 80. Götherström A, Anderung C, Hellborg L, Elburg R, Smith C, Bradley DG, Ellegren H: **Cattle domestication in the Near East was followed by hybridization with aurochs bull in Europe.** *Proc Biol Sci* 2005, **272**:2345–2350. *Erratum in: Proc Biol Sci* 2005, **272**:2660.
 81. Bollongino R, Elsner J, Vigne JD, Burger J: **Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle.** *PLoS One* 2008, **3**:e3418.
 82. Svensson E, Götherström A: **Temporal fluctuations of Y-chromosomal variation in *Bos taurus*.** *Biol Lett* 2008, **4**:752–754.
 83. Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hänni C, Fortea J, de la Rasilla M, Bertranpetit J, Rosas A, Pääbo S: **The derived FOXP2 variant of modern humans was shared with Neandertals.** *Curr Biol* 2007, **17**:1908–1912.
 84. Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, Hughes D, Rohland N, Pili E, Longo L, Condemi S, de la Rasilla M, Fortea J, Rosas A, Stoneking M, Schöneberg T, Bertranpetit J, Hofreiter M: **A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals.** *Science* 2007, **318**:1453–1455.
 85. Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, Castañón P, Cieslak M, Lippold S, Llorente L, Malaspinas AS, Slatkin M, Hofreiter M: **Coat color variation at the beginning of the horse domestication.** *Science* 2009, **324**:485.
 86. Hauswirth WW, Dickel CD, Rowol DJ, Hauswirth MA: **Inter- and intra-population studies of ancient humans.** *Experientia* 1994, **50**:585–591.
 87. Marota I, Basile C, Ubaldi M, Rollo FD: **DNA decay rate in papyri and human remains from Egyptian archaeological sites.** *Am J Phys Anthropol* 2002, **117**:310–318.
 88. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S: **Ancient DNA.** *Nat Rev Genet* 2001B, **2**:353.
 89. Richards MB, Sykes BC, Hedges REM: **Authenticating DNA extracted from ancient skeletal remains.** *J Archaeol Sci* 1995, **22**:291–299.
 90. Handt O, Krings M, Ward RH, Pääbo S: **The retrieval of ancient human DNA sequences.** *Am J Hum Genet* 1996, **59**:368–376.
 91. Kolman C, Tuross N: **Ancient DNA analysis of human populations.** *Am J Phys Anthropol* 2000, **111**:5–23.
 92. Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, Penkman KEH, Prangenberg K, Nielsen-Marsh CM, Jans ME, Arthur P, Lynnerup N, Turner-Walker G, Biddle M, Kjølbye-Biddle B, Collins MJ: **Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy.** *J Archaeol Sci* 2005, **32**:785–793.
 93. Gilbert MT, Wilson AS, Bunce M, Hansen AJ, Willerslev E, Shapiro B, Higham TF, Richards MP, O'Connell TC, Tobin DJ, Janaway RC, Cooper A: **Ancient mitochondrial DNA from hair.** *Curr Biol* 2004, **14**:R463–R464.
 94. Salamon M, Tuross N, Arensburg B, Weiner S: **Relatively well preserved DNA is present in the crystal aggregates of fossil bones.** *Proc Natl Acad Sci USA* 2005, **102**:13783–13788.
 95. Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A: **Distribution patterns of post-mortem damage in human mitochondrial DNA.** *Am J Hum Genet* 2003, **72**:32–47.
 96. Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A: **Characterization of genetic miscoding lesions caused by postmortem damage.** *Am J Hum Genet* 2003, **72**:48–61.
 97. Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, Bertorelle G: **Evidence for a genetic discontinuity between Neandertals and 24000-year-old anatomically modern Europeans.** *Proc Natl Acad Sci USA* 2003, **100**:6593–6597.
 98. Birladen J, Gilbert MTP, Willerslev E: **800 000 year old mammoth DNA, modern elephant DNA or PCR artefact?** *Biol Lett* 2007, **3**:55–56.
 99. Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM: **Patterns of nuclear DNA degeneration over time, a case study in historic teeth samples.** *Mol Ecol* 2003, **12**:1087–1093.
 100. Malmström H, Stora J, Dalén L, Holmlund G, Götherström A: **Extensive human DNA contamination in extracts from ancient dog bones and teeth.** *Mol Biol Evol* 2005, **22**:2040–2047.
 101. Caramelli D, Milani L, Vai S, Modi A, Pecchioli E, Girardi M, Pili E, Lari M, Lippi B, Ronchitelli A, Mallegni F, Casoli A, Bertorelle G, Barbujani G: **A 28000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences.** *PLoS One* 2008, **3**:e2700.
 102. Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM: **New developments in ancient genomics.** *Trends Ecol Evol* 2008, **23**:386–393.
 103. Shendure J, Mitra RD, Varma C, Church GM: **Advanced sequencing technologies: methods and goals.** *Nat Rev Genet.* 2004 May, **5**(5):335–344.
 104. Margulies M, Egholm M, Altam WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376–380.
 105. Meyer M, Briggs AW, Maricic T, Höber B, Höffner B, Krause J, Weihmann A, Pääbo S, Hofreiter M: **From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing.** *Nucleic Acids Res* 2008, **36**:e5.
 106. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P: **Real-time DNA sequencing using detection of pyrophosphate release.** *Anal Biochem* 1996, **242**:84–89.
 107. Ronaghi M, Uhlén M, Nyrén P: **A sequencing method based on real-time pyrophosphate.** *Science* 1998, **281**:363–365.
 108. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphree RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Ramm P, Miller W, Schuster SC: **Metagenomics to paleogenomics: large-scale sequencing of mammoth.** *Science* 2006, **311**:392–394.
 109. Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, Tikhonov A, Raney B, Patterson N, Lindblad-Toh K, Lander ES, Knight JR, Irzyk GP, Fredrikson KM, Harkins TT, Sheridan S, Pringle T, Schuster SC: **Sequencing the nuclear genome of the extinct woolly mammoth.** *Nature* 2008, **456**:387–390.
 110. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S: **Genetic history of an archaic hominin group from Denisova Cave in Siberia.** *Nature* 2010, **468**:1053–1060.
 111. Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A, Bertranpetit J, Krause J: **Genetic characterization of the ABO blood group in Neanderthals.** *BMC Evol Biol* 2008, **8**:342.
 112. Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A: **Bitter taste perception in Neanderthal through the analysis of the TAS2R38 gene.** *Biol Lett* 2009, **5**:809–811.
 113. Lari M, Rizzi E, Milani L, Corti G, Balsamo C, Vai S, Catalano G, Pili E, Longo L, Condemi S, Giunti P, Hänni C, De Bellis G, Orlando L, Barbujani G, Caramelli D: **The microcephalin ancestral allele in a Neanderthal individual.** *PLoS One* 2010, **5**:e10648.
 114. Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, Soares P, Luciani S, Marota I, De Bellis G, Richards MB, Rollo F: **Complete mitochondrial genome sequence of the Tyrolean Iceman.** *Curr Biol* 2008, **18**:1687–1693.

115. Olivieri C, Ermini L, Rizzi E, Corti G, Bonnal R, Luciani S, Marota I, De Bellis G, Rollo F: **Characterization of nucleotide misincorporation patterns in the iceman's mitochondrial DNA.** *PLoS One* 2010, **5**:e8629.
116. Stiller M, Knapp M, Stenzel U, Hofreiter M, Meyer M: **Direct multiplex sequencing (DMPS) a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA.** *Genome Res* 2009, **19**:1843–1848.
117. Lari M, Rizzi E, Mona S, Corti G, Catalano G, Chen K, Vernesi C, Larson G, Boscato P, De Bellis G, Cooper A, Caramelli D, Bertorelle G: **The complete mitochondrial genome of an 11450-year-old auroch (*Bos primigenius*) from Central Italy.** *BMC Evol Biol* 2011, **11**:32.
118. Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev VB, Golovanova LV, de la Rasilla M, Fortea J, Rosas A, Pääbo S: **Target retrieval and analysis of five Neandertal mtDNA genomes.** *Science* 2009, **325**:318–321.
119. Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, Pääbo S: **The complete mitochondrial DNA genome of an unknown hominin from southern Siberia.** *Nature* 2010, **464**:894–897.
120. Maricic T, Pääbo S: **Optimization of 454 sequencing library preparation from small amounts of DNA permits sequence determination of both DNA strands.** *Biotechniques* 2009, **46**:51–57.
121. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: **Exome sequencing identifies the cause of a Mendelian disorder.** *Nat Genet* 2010, **42**:30–35.
122. Choi M, Scholl UJ, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP: **Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.** *Proc Natl Acad Sci USA* 2009, **106**:19096–20101.
123. Bau S, Schracke N, Kränzle M, Wu H, Stähler PF, Hoheisel JD, Beier M, Summerer D: **Target next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluid DNA array.** *Anal Bioanal Chem* 2009, **393**:171–175.
124. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PL, Xuan Z, Rooks M, Bhattacharjee A, Brizuela L, Albert FW, de la Rasilla M, Fortea J, Rosas A, Lachmann M, Hannon GJ, Pääbo S: **Targeted investigation of the Neanderthal genome by array-based sequence capture.** *Science* 2010, **328**:723–725.
125. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nussbaum C, Gnirke A: **Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.** *Genome Biol* 2011, **12**:R18.
126. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Dettler JC, Pääbo S, Rubin EM: **Genomic sequencing of Pleistocene cave bears.** *Science* 2005, **309**:597–599.
127. Briggs AW, Stenzel U, Johnson P, Green R, Kelso J, Prüfer K, Meyer M, Krause J, Ronan M, Lachmann M, Pääbo S: **Patterns of damage in genomic DNA sequences from a Neandertal.** *Proc Natl Acad Sci USA* 2007, **104**:14616–14621.
128. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L: **mapDamage: testing for damage patterns in ancient DNA sequences.** *Bioinformatics* 2011, **27**:2153–2155.
129. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Pääbo S: **Analysis of one million base pairs of Neandertal DNA.** *Nature* 2006, **444**:330–336.
130. Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM: **Sequencing and analysis of Neandertal genomic DNA.** *Science* 2006, **314**:1113–1118.
131. Wall JD, Kim SK: **Inconsistencies in Neandertal genomic DNA sequences.** *PLoS Genet* 2007, **3**:1862–1866.
132. Lalueza-Fox C, Gilbert T: **Palaeogenomics of archaic hominins.** *Curr Biol* 2011, **21**:R1002–R1009.
133. Malmström H, Gilbert MT, Thomas MG, Brandström M, Storå J, Molnar P, Andersen PK, Bendixen C, Holmlund G, Götherström A, Willerslev E: **Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians.** *Curr Biol* 2009, **19**:1758–1762.
134. Gilbert MT, Kivisild T, Grönnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF, Rasmussen M, Metspalu M, Higham TF, Schwenninger JL, Nathan R, De Hoog CJ, Koch A, Møller LN, Andreasen C, Meldgaard M, Vilems R, Bendixen C, Willerslev E: **Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland.** *Science* 2008, **320**:1787–1789.
135. Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PG, Germonpré M, Götherström A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC: **Whole-genome shotgun sequencing of mitochondrial from ancient hair shafts.** *Science* 2007, **317**:1927–1930.
136. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Birladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grönnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Vilems R, Nielsen R, Krogh A, Wang J, Willerslev E: **Ancient human genome sequence of an extinct Palaeo-Eskimo.** *Nature* 2010, **463**:757–762.
137. Allentoft M, Schuster SC, Holdaway R, Hale M, McLay E, Oskam C, Gilbert MT, Spencer P, Willerslev E, Bunce M: **Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data.** *Biotechniques* 2009, **46**:195–200.
138. Miller W, Drautz DI, Janecka JE, Lesk AM, Ratan A, Tomsho LP, Packard M, Zhang Y, McClellan LR, Qi J, Zhao F, Gilbert MT, Dalén L, Arsuaga JL, Ericson PG, Huson DH, Helgen KM, Murphy WJ, Götherström A, Schuster SC: **The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*).** *Genome Res* 2009, **19**:213–220.
139. Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, Steinmann KE, Kapranov P, Thompson JF, Zazula G, Froese D, Moltke I, Shapiro B, Hofreiter M, Al-Rasheid KA, Gilbert MT, Willerslev E: **True single-molecule DNA sequencing of a pleistocene horse bone.** *Genome Res* 2011, **21**:1705–1719.
140. Ho SY, Gilbert MT: **Ancient mitogenomics.** *Mitochondrion* 2010, **10**:1–11.
141. Stoneking M, Krause J: **Learning about human population history from ancient and modern genomes.** *Nat Rev Genet* 2011, **12**:603–614.

doi:10.1186/1297-9686-44-21

Cite this article as: Rizzi et al.: Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution* 2012 **44**:21.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

